

Analysis of Building Energy Efficiency Dataset using Naive Bayes Classification Classifier

by Budi Prasetyo

Submission date: 23-May-2019 02:53PM (UTC+0700)

Submission ID: 1134793484

File name: sis-of-Building-Energy-fullpaper-20190523-073609-ZaPFz82zww.docx (53.63K)

Word count: 1629

Character count: 8371

Analysis of Building Energy Efficiency Dataset using Naive Bayes Classification Classifier

Budi Prasetyo¹, Alamsyah², Much Aziz Muslim³

^{1,2,3}Department of Computer Science, FMIPA, Universitas Negeri Semarang

¹budipras@mail.unnes.ac.id, ²alamsyah@mail.unnes.ac.id

Abstract. The phenomenon of global warming has become a very important and serious issue in recent years. One effort to reduce emissions by applying the green architecture concept to buildings. The purpose of this paper is to classify energy efficiency data sets. The method used by the Naïve Bayesian Classifier algorithm. Results obtained by Kappa statistics Y1 (0.3357) and Y2 (0.4203); Mean absolute error Y1 (0,0368), Y2 (0,0335); Root mean squared error Y1(0,1491), Y2(0,1412). The results of the analysis showed overall height, relative compactness, wall area, and glazing area affected the heating and cooling loads, while the surface area affected the heating load and the roof area affected the cooling load.

1. Introduction

The phenomenon of global warming has become a very important and serious issue in the last few years, this is not separated from the increasingly negative impact of global warming. The impact of global warming is felt by both developed and developing countries. Not only among the government, or diplomacy between countries in international forums, but also among environmental activists and even, news about the effects of global warming has always been a hot conversation in various media and the general public. The United Nations Inter-governmental Panel on Climate Change states that the main cause of global warming is caused by greenhouse gases due to human activities. The composition of greenhouse gases is 73% methane, 6% nitrogen oxide and 5% fluorocarbons, and the largest is 76% CO2 carbon dioxide. The amount of CO2 is a significant factor to increase global warming [1]. High CO2 emissions are generated from human activities (anthropogenic) [2], one of which is in buildings. CO2 emissions were produced both during the process of making building materials, the design of house construction to household activities when the house was occupied [3].

The building sector turned out to be the biggest factor causing global warming. The International Energy Agency released a publication that estimates that buildings are responsible for more than 40% of the world's total energy consumption and 24% of contributors to CO2 emissions globally. Buildings consume more than a third of the world's energy compared to the industrial and transportation sectors in consuming fossil energy [4]. As the urban population increases and the population explodes, energy consumption is projected to continue to rise in the building sector, especially in developing countries [5]. Green Building which in this case is also referred to as green architecture (Green Architecture) is an issue that starts to emerge following the issue of global warming [6]. With the concept of green building, it is expected to be able to reduce energy use efficiently and the impact of pollution while building design becomes environmentally friendly. This makes the building-saving classification model very important for every building developer. So that a certain method is needed to classify energy-efficient building models through recording existing data. Classification problems can be done using methods in data mining [7].

Data mining is an activity to explore and obtain information from large amounts of data [8]. Data mining is used to analyze and extract knowledge automatically. One of the data mining activities is classification techniques that for predicting a value from the target variable category (label category for each data [9]. Activities in classification are extracting data and then predicting the main classification is to predict how the classification value of the data [10]. To do the classification process a classification algorithm is needed. There are several classification algorithms in data mining that are quite popular, one of which is the Naïve Bayesian Classifier algorithm [11]. Naïve Bayesian Classifier is one classification algorithm that works using a statistical (probability) approach. Naïve Bayesian is classified as a simple, easy to use classification algorithm where each attribute is independent and gives the possibility (probability) of the final decision [12]. The results obtained from the classification process need to be measured for performance to test the level of accuracy. In the world of data mining, algorithm performance can be measured by Confusion Matrix. This is a method usually used to calculate accuracy. Confusion Matrix is also known as a contingency table or error matrix [13]. This calculation produces 4 outputs, namely: recall, precision, accuracy and error rate. In addition to evaluation in a calculation, it also needs to be evaluated graphically. Graphical performance measurements on an algorithm can be done using the Receiver Operating Characteristic (ROC) curve or ROC curve. The method used is by calculating the area under the curve called AUC (Area Under the ROC Curve) [11]. The ROC curve is very useful in dealing with classification problems because it graphically shows the visualization of its performance [14].

2. Analysis and Data Pre-processing

At this stage data analysis will be used, whether it contains noise or is ready to use. The data used in this study uses the Energy efficiency data set dataset, which is taken from the UCI Machine Learning Repository (Tsanas, 2014). The information about the data is

- a. Many samples: 768 samples of building shapes.
- b. Number of Attributes: 8, each according to the parameters in the building.
- c. Attribute information:
 - X1 relative compactness (relative compactness)
 - X2 surface area (surface area)
 - X3 wall area (wall area)
 - X4 roof area wide roof
 - X5 overall height (overall height)
 - X6 orientation (orientation)
 - X7 glazing area (glass area)
 - X8 glazing area distribution (glass location distribution)
 - y1 heating load (heating load)
 - y2 cooling load (cooling load)
- d. Attribute value missing: none

At the preprocessing stage the data is broken down into two parts, namely data for training and data for testing.

3. Result

To carry out classification, the data used in this study uses the Energy Efficiency dataset, which is taken from the UCI Machine Learning Repository. From the dataset, there are 768 data, sample data as shown in Table 1.

X2	X3	X4	X5	X6	X7	X8	Y1	Y2
514,50	294,00	110,25	7,00	2	0,00	0	15,55	21,33
514,50	294,00	110,25	7,00	3	0,00	0	15,55	21,33
514,50	294,00	110,25	7,00	4	0,00	0	15,55	21,33
514,50	294,00	110,25	7,00	5	0,00	0	15,55	21,33
563,50	318,50	122,50	7,00	2	0,00	0	20,84	28,28

563,50	318,50	122,50	7,00	3	0,00	0	21,46	25,38
563,50	318,50	122,50	7,00	4	0,00	0	20,71	25,16
563,50	318,50	122,50	7,00	5	0,00	0	19,68	29,60
588,00	294,00	147,00	7,00	2	0,00	0	19,50	27,30
588,00	294,00	147,00	7,00	3	0,00	0	19,95	21,97
588,00	294,00	147,00	7,00	4	0,00	0	19,34	23,49
588,00	294,00	147,00	7,00	5	0,00	0	18,31	27,87
612,50	318,50	147,00	7,00	2	0,00	0	17,05	23,77
612,50	318,50	147,00	7,00	3	0,00	0	17,41	21,46
661,50	416,50	122,50	7,00	4	0,00	0	24,77	29,79
661,50	416,50	122,50	7,00	5	0,00	0	23,93	29,40
686,00	245,00	220,50	3,50	2	0,00	0	6,07	10,90
686,00	245,00	220,50	3,50	3	0,00	0	6,05	11,19
686,00	245,00	220,50	3,50	4	0,00	0	6,01	10,94
...

The classification modeling results using Naïve Bayesian then performed accuracy. The results of performance accuracy are shown in the Table 2.

5 Table 2. Results of Performance classification

	true 2	true 3	true 1	true 4	true 5	class precision
pred. 2	360	0	0	0	0	100.00%
pred. 3	13	147	0	80	0	61.25%
pred. 1	4	0	20	0	0	83.33%
pred. 4	0	23	0	78	3	75.00%
pred. 5	0	0	0	9	31	77.50%
class recall	95.49%	86.47%	100.00%	46.71%	91.18%	
accuracy: 82.81% +/- 2.56%						

The results of the Naïve Bayesian Classifier algorithm produce kappa statistics Y1 (0.3357) and Y2 (0.4203); Mean absolute error Y1 (0.0368), Y2 (0.0335); Root mean squared error Y1 (0.1491), Y2 (0.1412).

4. Conclusion

Development of a green architecture model is carried out by conducting accuracy testing using Rapid miner software. Furthermore, software needs to be made with certain programming languages. The classification results show the accuracy value of 82.81%, meaning that it is quite high and with an average class recall of 83.97% and an average class precision of 79.42%. The building classification results consist of 5 classifications in heating and cooling loads. The results of the Naïve Bayesian Classifier algorithm produce kappa statistics Y1 (0.3357) and Y2 (0.4203); Mean absolute error Y1 (0.0368), Y2 (0.0335); Root mean squared error Y1 (0.1491), Y2 (0.1412).

References

- [1] Sagheb A, Ehsan V, and Pradeep K R 2011 *Natl. Conf. Recent Trends Civ. Mech. Eng., Int. Inst. Inf. Technol.*
- [2] Fujita Y, Hiroshi M, and Ho C S 2009 *Int. J. Low-Carbon Technol* 4 16
- [3] Sudjon P, and Chendy O Y 2011 *J. Tek. Lingkungan* 17 2
- [4] Roaf S, David C, and Fergus N 2005 *Adapting Buildings and Cities for Climate Change*. (USA: Architectural Press)
- [5] Odebiyi S O, Suneetha S A K, and Braimoh 2020 *J. Altern. Perspect. Soc. Sci.* 2 2
- [6] Vezolli and Manzini 2008 *Design for Environmental Sustainability*. (London: Springer)
- [7] Martin A, T Miranda L, and V Prasanna V 2012 *Int. Conf. Emerg. Trends Sci. Eng. Technol. (INCOSSET)*
- [8] Olson D L, and Yong S 2007 *Introduction to Business Data Mining* (New York: McGraw-Hill)
- [9] Han J, Micheline K and Jian P 2012 *Data Mining: Concepts and Techniques, Third Edition* (USA:Elsevier)
- [10] Thabtah F, Hammoud S, and Abdeljaber Hussein 2015 *J. Parallel Process. Lett.* 25 2
- [11] Gorunescu, F 2011 *Data Mining Concept Model and Techniques* (Berlin: Springer)
- [12] Xhemali, D, Hinde C J, and Stone R G 2009 *Int. J. Comput. Sci. Issues* 4 1
- [13] Stehman, S V 1997 *Remote Sens. Environ.* 62 1
- [14] Fawcett, T 2006 *Pattern Recognit. Lett.* 27
- [15] Tsanas A, Angeliki X 2012 *Energy Build.* 49

Analysis of Building Energy Efficiency Dataset using Naive Bayes Classification Classifier

ORIGINALITY REPORT

10%

SIMILARITY INDEX

4%

INTERNET SOURCES

4%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

- 1** Submitted to Universiti Teknologi MARA
Student Paper 2%
- 2** Submitted to University of Southern California
Student Paper 2%
- 3** Tarek Samarji, Adnan Jouni, Ali Karaki. "Net zero energy buildings: Application in Lebanon on a typical residential building", 2012 International Conference on Renewable Energies for Developing Countries (REDEC), 2012
Publication 2%
- 4** TOPRAK, Ahmet, KOKLU, Nigmet, TOPRAK, Aysegul and OZCAN, Recai. "Comparison of Classification Techniques on Energy Efficiency Dataset", Selçuk Üniversitesi Teknoloji Fak., 2017.
Publication 1%
- 5** eswo.org
Internet Source 1%

6	mizar.uwb.edu.pl Internet Source	1 %
7	www.researchgate.net Internet Source	1 %
8	Submitted to Eastern Mediterranean University Student Paper	1 %
9	Submitted to University of Portsmouth Student Paper	1 %
10	Grant McDougal, Rajiv Agarwal. "Clinical performance characteristics of hemodialysis graft monitoring", Kidney International, 2001 Publication	<1 %
11	Submitted to University of Leeds Student Paper	<1 %

Exclude quotes On

Exclude bibliography On

Exclude matches < 4 words